REMARKS

Claims 1-15, and 17-28 are currently pending in the subject application and are presently under consideration. Claims 1, 19, 22, 24 and 26 have been amended and claim 16 has been cancelled herein, as shown on pages 2-5 of the Reply. The below comments present in greater detail distinctive features of applicants' claimed invention over the cited art that were conveyed to the Examiner over the telephone on December 20, 2007.

Favorable reconsideration of the subject patent application is respectfully requested in view of the comments and amendments herein.

I.   Rejection of Claims 26-28 Under 35 U.S.C. §102(e)

Claims 26-28 stand rejected under 35 U.S.C. §102(e) as being anticipated by Harik (US 7,213,198). Withdrawal of this rejection is requested for the following reasons. The cited reference fails to disclose or suggest all aspects set forth in the subject claims.

> A single prior art reference anticipates a patent claim only if it *expressly or inherently describes each and every limitation set forth in the patent claim*. *Trintec Industries, Inc. v. Top-U.S.A. Corp.*, 295 F.3d 1292, 63 USPQ2d 1597 (Fed. Cir. 2002); *See Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987). The *identical invention must be shown in as complete detail as is contained in the ... claim*. *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989) (emphasis added).

The claimed invention relates to a system and method of facilitating incremental web crawls using chunks. To this end, amended independent claim 26 recites *means for placing items with similar properties into respective chunks and means for storing at least some of the properties associated with the respective chunk, wherein the properties are at least one of average time between change or average importance of documents comprising a particular chunk, and employing them to facilitate an incremental web re-crawl.* Harik is silent regarding such novel features.

Harik relates to grouping web pages from a search and techniques for displaying similar topics that are linked to the web pages. At the cited portions, Harik discloses a system that groups web pages by topic. A hash table is utilized to organize the relationship between web

6

pages. The hash table contains properties of the individual documents, which are analyzed and used to relate each pair of expanded search set pages of a selected document. Thus, the system employs the properties a document and its links in the table for a further search. In contrast, the claimed invention stores properties of a particular chunk, which are the averaged properties of all the documents such as average time of change or average importance, and this averaged property of the chunk are employed to facilitate a re-crawl. Thus, Harik is silent regarding *means for storing at least some of the properties associated with the respective chunk, wherein the properties are at least one of average time between change or average importance of documents comprising a particular chunk, and employing the properties to facilitate an incremental web re-crawl* as recited by the subject claim. Hence, it can be concluded that Harik does not teach an identical invention as recited in the subject claims.

Accordingly, it is requested that this rejection with respect to independent claim 26 should be withdrawn.

II.   Rejection of Claims 1-4, 6-14, 17 and 19-21 Under 35 U.S.C. §103(a)

Claims 1-4, 6-14, 17 and 19-21 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Harik (US 7,213,198) in view of Agarwal, *et al.* (US 2004/0225963). Withdrawal of this rejection is requested for the following reasons. Harik and Agarwal *et al.*, alone or in combination, fail to teach or suggest all aspects set forth in the subject claims.

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Information gathered from a web crawl is indexed and chunked based on similar properties like average time between change and average importance of the retrieved documents. This chunk map then is employed to determine which chunks should be re-crawled. To this end, amended independent claim 1 recites *a system that facilitates incremental web crawls comprising: **an indexer that places items with similar properties into respective chunks;** and a chunk map that stores at least some of the properties associated with the respective chunk, **wherein the properties are at least one of average time between change or average importance of documents comprising a particular chunk, the chunk map employed to facilitate an incremental web re-crawl.*** Independent claim 19 recites similar features. Harik and Agarwal *et al.*, are silent regarding such novel features.

Harik relates to grouping web pages from a search and techniques for displaying similar topics that are linked to the web pages. At the cited portions, Harik discloses a system that groups web pages by topic. Forward links and backward links to the web pages are grouped as expanded groups and displayed, grouped separated graphically on the screen. However, Harik is silent regarding *a chunk map that stores at least some of the properties associated with the respective chunk,* **wherein the at least one of the properties are average time between change and average importance of documents.** Further, at the cited portions, Harik discloses a hash table utilized to organize the relationship between the web pages and their back links. The hash table is used for more efficiently grouping the links to a web page, and not to facilitate a re-crawl. The hash table contains properties of the individual documents, which are analyzed and used to relate each pair of expanded search set pages of a selected document. In contrast, the claimed invention stores properties of a particular chunk, which are the averaged properties of all the documents such as average time of change or average importance, and this averaged property of the chunk are employed to facilitate a re-crawl. Thus, Harik is silent regarding **wherein the properties are at least one of average time between change or average importance of documents comprising a particular chunk, the chunk map employed to facilitate an incremental web re-crawl.**

Agarwal *et al.* relates to an efficient technique for indexing document repositories based on landmark entries. Documents are divided into a plurality of blocks, each block is associated with a landmark entry and the document's position relative to the landmark entry is recorded. The document entries and landmark entries are included in the repository index. At the cited portions, Agarwal *et al.* discloses a search engine that utilizes a crawler that automatically finds documents in the document repository and updates the search engine's records. The document repository represents a collection of all information accessible through the computer network. In contrast, applicants claimed invention teaches an indexer that places items with similar properties into chunks, and a map that stores the chunk level properties like average time between change and average importance of the items, in a chunk map. This chunk map is employed to facilitate an incremental web re-crawl. Thus, Agarwal *et al.* is silent regarding **the chunk map employed to facilitate an incremental web re-crawl.** By chunking the documents on the basis of properties like time of change or importance, and storing these in a chunk map, the crawler does not need to do any per document analysis to determine the frequency of when a

document should be crawled. Because a document is in a chunk and the chunk map has average properties and is identified ad the chunk to be re-crawled implies that the document should be crawled. This allows for faster processing of documents by the crawler.

In view of at least the foregoing it is readily apparent that Harik and Agarwal *et al.* either alone or in combination do not teach or suggest each and every element set forth in the applicants' subject claims. Accordingly it is requested that this rejection should be withdrawn.

III. Rejection of Claim 5 Under 35 U.S.C. §103(a)

Claim 5 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Harik in view of Agarwal, *et al.* and further in view of Eichstaedt, *et al.* (US 6,182,085). It is respectfully requested that this rejection be withdrawn for at least the following reasons. Harik, Agarwal *et al.* and Eichstaedt *et al.* do not teach or suggest all aspects set forth in the subject claims. Claim 5 depends from independent claim 1, and as discussed *supra*, Harik and Agarwal *et al.*, alone or in combination, do not disclose all features of independent claim 1. Eichstaedt *et al* relates, and does not make up for the aforementioned deficiencies of Harik and Agarwal *et al.* with respect to independent claim 1. Thus, the subject invention as recited in the subject claims is not obvious over the combination of Harik and Agarwal *et al.* and Eichstaedt *et al.* Accordingly, it is respectfully submitted that this rejection of claim 5 should be withdrawn.

IV. Rejection of Claims 15-16 Under 35 U.S.C. §103(a)

Claims 15-16 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Harik in view of Agarwal, *et al.* and further in view of Najork, *et al.* (US 6,263,364). It is respectfully requested that this rejection be withdrawn for at least the following reasons. Harik, Agarwal *et al.* and Najork *et al.* do not teach or suggest all aspects set forth in the subject claims. Claims 15 and 16 depend from independent claim 1, and as discussed *supra*, Harik and Agarwal *et al.*, alone or in combination, do not disclose all features of independent claim 1. Najork *et al* relates to a web crawler that downloads documents from a plurality of host computers and enqueues document addresses, where each queue has documents sharing a common host component, and does not make up for the aforementioned deficiencies of Harik and Agarwal *et al.* with respect to independent claim 1. Thus, the subject invention as recited in the subject claims is not obvious over the combination of Harik and Agarwal *et al.* and Najork *et al.* Accordingly, it is

respectfully submitted that this rejection of claims 15 and 16 should be withdrawn.

V.   Rejection of Claim 18 Under 35 U.S.C. §103(a)

Claim 18 stands rejected under 35 U.S.C. §103(a) as being unpatentable over Harik in view of Agarwal, *et al.* and further in view of Acharaya, *et al.* (US 2007/0094255). It is respectfully requested that this rejection be withdrawn for at least the following reasons. Harik, Agarwal *et al.* and Acharaya *et al.* do not teach or suggest all aspects set forth in the subject claims. Claim 18 depends from independent claim 1, and as discussed *supra*, Harik and Agarwal *et al.*, alone or in combination, do not disclose all features of independent claim 1. Acharaya *et al* relates to document scoring based on link-based criteria, and does not make up for the aforementioned deficiencies of Harik and Agarwal *et al.* with respect to independent claim 1. Thus, the subject invention as recited in the subject claims is not obvious over the combination of Harik and Agarwal *et al.* and Acharaya *et al.* Accordingly, it is respectfully submitted that this rejection of claim 18 should be withdrawn.

VI.   Rejection of Claims 22-23 Under 35 U.S.C. §103(a)

Claims 22-23 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Harik in view of Najork, *et al.* (US 6,263,364). Withdrawal of this rejection is requested for the following reasons. Harik and Najork *et al.*, alone or in combination, fail to teach or suggest all aspects set forth in the subject claims.

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Amended independent claim 22 recites similar features as independent claim 1, namely *accessing **a chunk map containing properties** associated with respective chunks of data as a result of one or more web crawls, **wherein the at least one of the properties are average time of change or average importance of documents; and periodically determining, based on the properties in the chunk map, whether to re-crawl one or more of the chunks of data.*** Harik and Najork *et al.*, alone or in combination, fail to disclose such novel features.

As discussed *supra*, Harik fails to disclose *a **chunk map containing properties** associated with respective chunks of data as a result of one or more web crawls, **wherein the at***

least one of the properties are average time of change or average importance of documents as recited by the subject claims. Najork *et al* relates to a web crawler that downloads documents from a plurality of host computers and enqueues document addresses, where each queue has documents sharing a common host component. At page 12 of the Office Action, the Examiner contends that Najork *et al.* teaches novel features of the claimed invention. Applicants' representative avers to the contrary. At the cited portions, Najork *et al.* discloses assigning a processed document to a priority level subqueue, based on a priority level which factors properties including the document's expiration date, rate of change and the host computer of its URL. Thus, Najork *et al.* discloses a system that assigns a priority level to a document based upon its properties, and queues documents having a same priority level to be crawled. In contrast, the claimed invention chunks documents having similar properties into chunks and employs a chunk map that has properties which are averaged properties of all the documents, *to periodically determine, based on the properties in the chunk map* to perform a re-crawl. Thus, Najork *et al.* does not disclose *periodically determining, based on the properties in the chunk map, whether to re-crawl one or more of the chunks of data* as recited by the subject claims.

In view of at least the foregoing it is readily apparent that Harik and Najork *et al.* either alone or in combination do not teach or suggest each and every element set forth in the applicants' subject claims. Accordingly it is requested that this rejection should be withdrawn.

VII.   Rejection of Claims 24-25 Under 35 U.S.C. §103(a)

Claims 24-25 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Harik in view of Kosiba, *et al.* (US 2003/0221014). Withdrawal of this rejection is requested for the following reasons. Harik and Kosiba *et al.*, alone or in combination, fail to teach or suggest all aspects set forth in the subject claims.

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Amended independent claim 24 recites similar features as independent claim 1, namely *wherein the average of the at least one of the properties of all the document files determines if the document should be re-crawled.* As discussed *supra*, Harik does not teach all the features of independent claim 1. Kosiba *et al.* relates to a method for guaranteed delivery of multimedia content based on terminal capabilities, and does not compensate for the aforementioned deficiencies of Harik regarding amended

independent claim 24. In view of at least the foregoing it is apparent that Harik and Kosiba et *al.* either alone or in combination do not teach or suggest each and every element set forth in the applicants' subject claims. Accordingly it is requested that this rejection should be withdrawn.

<div align="center">CONCLUSION</div>

The present application is believed to be in condition for allowance in view of the above comments and amendments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP511US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicants' undersigned representative at the telephone number below.

Respectfully submitted,

AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/
Himanshu S. Amin
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP
24TH Floor, National City Center
1900 E. 9TH Street
Cleveland, Ohio 44114
Telephone (216) 696-8730
Facsimile (216) 696-8731